



HAL
open science

High-throughput functional profiling of the human fungal pathogen *Candida albicans* genome

Murielle Chauvel, Sophie Bachellier-Bassi, Anne-Marie Guérout, Keunsook K Lee, Corinne Maufrais, Emmanuelle Permal, Juliana Pipoli da Fonseca, Sadri Znaidi, Didier Mazel, Carol A Munro, et al.

► **To cite this version:**

Murielle Chauvel, Sophie Bachellier-Bassi, Anne-Marie Guérout, Keunsook K Lee, Corinne Maufrais, et al.. High-throughput functional profiling of the human fungal pathogen *Candida albicans* genome. *Research in Microbiology*, 2023, 174 (3), pp.104025. 10.1016/j.resmic.2022.104025 . pasteur-04131602

HAL Id: pasteur-04131602

<https://pasteur.hal.science/pasteur-04131602v1>

Submitted on 16 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



ELSEVIER

Contents lists available at ScienceDirect

Research in Microbiology

journal homepage: www.elsevier.com/locate/resmic

Original Article

High-throughput functional profiling of the human fungal pathogen *Candida albicans* genome



Murielle Chauvel^a, Sophie Bachellier-Bassi^a, Anne-Marie Guérout^{b,1}, Keunsook K. Lee^{c,f}, Corinne Maufrais^{a,d}, Emmanuelle Permal^{a,b,d}, Juliana Pipoli Da Fonseca^e, Sadri Znaidi^{a,g}, Didier Mazel^b, Carol A. Munro^c, Christophe d'Enfert^{a,**}, Melanie Legrand^{a,*}

^a Institut Pasteur, Université Paris Cité, INRAE USC2019, Unité Biologie et Pathogénicité Fongiques, F-75015 Paris, France

^b Institut Pasteur, Université Paris Cité, UMR3525 CNRS, Unité Plasticité du Génome Bactérien, F-75015 Paris, France

^c Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, UK

^d Institut Pasteur, Université Paris Cité, Hub de Bioinformatique, F-75015 Paris, France

^e Institut Pasteur, Université Paris Cité, Plate-forme Technologique Biomics, Centre de Ressources et Recherches Technologiques (C2RT), F-75015 Paris, France

^f NGeneBio, 307 Daerung Post-tower 1, 288 Digital-ro, Guro-gu, Seoul 08390, Republic of Korea

^g Laboratoire de Microbiologie Moléculaire, Vaccinologie et Développement Biotechnologique, Institut Pasteur de Tunis, University of Tunis El Manar, Tunis-Belvédère, Tunisia

ARTICLE INFO

Article history:

Received 20 July 2022

Accepted 26 December 2022

Available online 30 December 2022

Keywords:

Candida albicans

ORFeome

Overexpression mutants

Bar-seq

Large-scale systematic studies

ABSTRACT

Candida albicans is a major fungal pathogen of humans. Although its genome has been sequenced more than two decades ago, there are still over 4300 uncharacterized *C. albicans* genes. We previously generated an ORFeome as well as a collection of destination vectors to facilitate overexpression of *C. albicans* ORFs. Here, we report the construction of ~2500 overexpression mutants and their evaluation by in vitro spotting on rich medium and in a liquid pool experiment in rich medium, allowing the identification of genes whose overexpression has a fitness cost. The candidates were further validated at the individual strain level. This new resource allows large-scale screens in different growth conditions to be performed routinely. Altogether, based on the concept of identifying functionally related genes by cluster analysis, the availability of this overexpression mutant collection will facilitate the characterization of gene functions in *C. albicans*.

© 2023 The Authors. Published by Elsevier Masson SAS on behalf of Institut Pasteur. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author. Fungal Biology and Pathogenicity Unit–INRAE USC2019, Mycology Department, Institut Pasteur, 25-28, rue du Docteur Roux, 75015 Paris, France.

** Corresponding author. Fungal Biology and Pathogenicity Unit–INRAE USC2019, Mycology Department, Institut Pasteur, 25-28, rue du Docteur Roux, 75015 Paris, France.

E-mail addresses: murielle.chauvel@pasteur.fr (M. Chauvel), sophie.bachellier-bassi@pasteur.fr (S. Bachellier-Bassi), amguerout@gmail.com (A.-M. Guérout), ks.kathy.lee@ngenebio.com (K.K. Lee), corinne.maufrais@pasteur.fr (C. Maufrais), emmanuelle.permal@pasteur.fr (E. Permal), juliana.pipoli-da-fonseca@pasteur.fr (J.P. Da Fonseca), sadri.znaidi@pasteur.fr (S. Znaidi), didier.mazel@pasteur.fr (D. Mazel), c.a.munro@abdn.ac.uk (C.A. Munro), christophe.denfert@pasteur.fr (C. d'Enfert), mlegrand@pasteur.fr (M. Legrand).

¹ Retired author.

<https://doi.org/10.1016/j.resmic.2022.104025>

0923-2508/© 2023 The Authors. Published by Elsevier Masson SAS on behalf of Institut Pasteur. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Candida albicans is a major fungal pathogen of humans. Despite available antifungal treatments, the mortality rate of invasive or disseminated candidiasis remains high [1,2]. In order to improve the management of patients suffering from *C. albicans* infections and their therapeutic outcome, new antifungal agents with improved efficacy are urgently needed. A complete functional annotation of the *C. albicans* genome is still lacking and 69% of the predicted open reading frames (ORFs) remain uncharacterized [3]. Novel whole-genome approaches would be a game changer to systematically determine the function of each ORF, better understand the pathobiology of this opportunistic pathogen and identify novel antifungals targets.

In this respect, ORFeomes that refer to the complete set of protein-encoding open reading frames (ORFs) in a genome allow

such functional studies of the corresponding proteomes. The *C. albicans* ORFeome has been previously established, covering 83% of its genome and constitutes a novel and important resource to implement systems biology analysis of *C. albicans* [4]. The *C. albicans* ORFeome has been exploited to generate collections of overexpression (OE) mutants that have been characterized individually, allowing the identification of regulators of morphogenesis fitness [5], and genome stability [6] with the recent identification of a novel regulator of chromosome stability [7]. In the genome-wide era, characterization of individual mutants remains a time-consuming and laborious process that can be alleviated by Signature-tagged mutagenesis (STM) [8,9]. STM is a high-throughput method in which each mutant is uniquely labeled by different sequence signature tags. These DNA signature tags, or molecular barcodes, allow the detection and quantification of individual mutants in a pool after exposure to any experimental condition, e.g., upon exposure to antifungal compounds. Several methods can be used for tag detection, such as barcode microarrays and next-generation DNA sequencing technologies (Bar-seq [10]). Microarray profiling has been combined to gene OE to identify genes whose OE affects biofilm formation [11] or colonization in the murine gut by *C. albicans* [12]. Bar-seq has also been combined with a partial ORFeome (572 genes) to identify mediators of fluconazole tolerance in *C. albicans* [13]. In addition to its increased robustness as compared to microarrays, one of the advantages of Bar-seq is the possibility to design multiplex experiments [10].

Here, we report the construction of (i) a collection of 2496 P_{TET} -driven OE plasmids, each containing a *C. albicans* ORF associated to a unique 25bp barcode and (ii) a collection of 2454 *C. albicans* OE mutant strains. As a proof of concept and because morphogenesis is a virulence factor of *C. albicans*, we exploited the collection to investigate the impact of OE of 2454 genes in *C. albicans* in terms of growth morphology. Also, fungal-specific genes whose inappropriate expression is toxic for the cell are potent targets to feed new antifungal drug research. Therefore, we performed an *in-vitro* screen for growth defects upon gene OE in pool experiments. We identified 193 genes whose controlled expression is required for optimal growth of *C. albicans* in rich medium.

We anticipate that the provision of a collection of barcoded OE plasmids, covering currently 40% of the *C. albicans* ORFeome, will stimulate genome-wide functional analysis, and in this way will improve our understanding of *C. albicans* biology.

2. Materials and methods

2.1. Strains and growth conditions

C. albicans strains used in this study are listed in Table 1. *C. albicans* strains were routinely cultured at 30 °C in YPD medium

(1% yeast extract, 2% peptone, 2% dextrose), or synthetic dextrose (SD) medium (0.67% yeast nitrogen base, 2% dextrose). *Escherichia coli* DH5 α or TOP10 strains were routinely cultured at 30 °C or 37 °C in LB or 2 YT supplemented with either 10 $\mu\text{g ml}^{-1}$ gentamicin or 50 $\mu\text{g ml}^{-1}$ kanamycin or 50 $\mu\text{g ml}^{-1}$ ticarcillin. Solid media were obtained by adding 2% agar.

2.2. Barcode pool design and amplification

A set of 240,000 orthogonal 25mer DNA barcode (BC) sequences designed as described previously [14] were filtered to exclude *AvrII* (605), *NheI* (1225), *KpnI* (1396) or *StuI* (1098) restriction sites. The resulting sequences (235,676) were BLASTed against the full-genome sequences of *C. albicans* (strain SC5314, assembly 21 [15]), *C. parapsilosis* (strain CDC317 [16]), *C. glabrata* (strain CBS138 [17]) and a set of sequences from our Cip-GTW vector series [4] to remove BC sequences that match genome/vector sequences with more than 50% identity (i.e. 13 nucleotides), yielding 65,777 BC sequences. The U1 (5'-GATGTCCACGAGGTCTCT-3') and D1 (5'-CGGTGTCGGTCTCGTAG-3') primers previously used for creating the barcoded *Saccharomyces cerevisiae* deletion collection [18] were added upstream (U1) and downstream (D1) of the BC sequences to allow for the amplification of the BC library. The resulting U1-BC-D1 sequences containing *StuI*, *NheI* or *KpnI* restriction sites (210, 99 and 93 sequences, respectively) were removed from the list. The remaining 65,375 U1-BC-D1 DNA sequences were synthesized on a glass slide using an array-based oligonucleotide library synthesis approach (SurePrint Oligonucleotide Libraries, array # 253796610002, Agilent Technologies). Recovery of the DNA oligonucleotides from the array surface was performed as described previously [19]. Briefly, 500 μl of nuclease-free water was pipetted onto a cover slide placed in an Agilent array hybridization chamber. The BC array was sandwiched on top of the cover slide and rotated overnight at 99 °C in a hybridization oven. The liberated oligonucleotides (in ~300 μl H₂O) were collected in a 1.5 ml Eppendorf tube and sonicated 3 times for 3–5 s at 50% amplitude using an MSE probe sonicator. The oligonucleotide pool concentration (3.5 $\mu\text{g ml}^{-1}$) was determined using a NanoVue Plus spectrophotometer. The BC pool was amplified by low cycle PCR using primers AVR1I-U1 (5'-ATCATCCTAGGGATGTCCACGAGGTCTCT-3'), and NHEI-D1 (5'-CTACTAGCTAGCCGGTGTCCGGTCTCGTAG-3'), introducing an *AvrII* and *NheI* restriction site, respectively (underlined) (all primer sequences are listed in Table 2). For the first PCR cycle, denaturation was performed at 94 °C for 1 min, followed by an annealing step at 55 °C for 20 s and an elongation step at 72 °C for 1 min. The following 25 cycles used 10, 20 and 10 s at each respective temperature, with a final elongation step of 5 min at 72 °C. The amplified BC library was aliquoted in PCR tubes and stored at –20 °C for subsequent directional cloning in Cip-GTW expression vectors using *AvrII* and *BmtI* restriction sites.

Table 1
Candida albicans strains used in this study.

Strain	Lab identifier	Parental strain	Relevant Genotype	Reference
SC5314	CEC1	–	Wild-type	Gillum (1984) [28]
SN76	CEC805	–	<i>ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/ iro1Δ::λimm434 arg4Δ/arg4Δ his1Δ/his1Δ</i>	Noble (2005) [22]
SN76 NIMX	CEC1447	CEC805	SN76 <i>ADH1/adh1::P_{TDH3-car}TA::SAT1</i>	Hernandez-Cervantes (2020)
SN76 ARG4	CEC4640	CEC1447	SN76 <i>ADH1/adh1::P_{TDH3-car}TA::SAT1 arg4Δ/CaARG4</i>	This study
SN76 ARG4 HIS1	CEC4642	CEC4640	SN76 <i>ADH1/adh1::P_{TDH3-car}TA::SAT1 arg4Δ/CaARG4 his1Δ::hisG/HIS1</i>	This study
SN76 ARG4 HIS1 URA3	CEC4664	CEC4642	SN76 <i>ADH1/adh1::PTDH3-carTA::SAT1 arg4Δ::/CaARG4 his1Δ::hisG/HIS1 RPS1/RPS1::Cip10</i>	This study

Table 2
Oligonucleotides used in this study.

Name	Sequence (5' to 3')
AVRII-U1	ATCATCCCTAGGGATGTCCACGAGGTCTCT
NHEI-D1	CTACTAGCTAGCCGGTGTCCGCTCTGCTAG
CipUL	ATACTACTGAAAATTCCTGACTTTC
CipUR	ATTACTATTACAATCAAAGGTGGTC
For barcode amplification	
Primer name	Primer sequence (5'-3')
Fd-Extension-0N-BC-SP1_Fwd	<u>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTAGGGATGTCCACGAGGTCT</u>
Fd-Extension-4N-BC-SP1_Fwd	<u>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCTAGGGATGTCCACGAGGTCT</u>
Fd-Extension-8N-BC-SP1_Fwd	<u>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNCTAGGGATGTCCACGAGGTCT</u>
Fd-Extension-12 N-BC-SP1_Fwd	<u>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNNNNCTAGGGATGTCCACGAGGTCT</u>
ReverseExtension-BC_Rev	<u>GTCCTGTTGGTCCGAGATGTGTATAAGAGACAGCTTTCAGTACAACGTCGAGCA</u>

Underlined sequences show the restriction sites.

Sequences in bold black show the Illumina overhangs.

Sequences in bold red show the spacer.

Underlined sequences show the DNA binding site.

2.3. Barcode cloning

To optimize barcode cloning in the Cip10- P_{TET} -GTW-SP vector [5], we used the phd/doc toxin-antitoxin modules [20]. The Doc toxin gene was inserted within the 17bp landing platform, leading to Cip10- P_{TET} -GTW-SP-DOC. In parallel, the chromosome of the *E. coli* recipient strain had to be modified to integrate the gene encoding the PHD antitoxin at *attTn7* (*glmS::PHD*), yielding strain ECC1081 that allows propagation of the toxin-bearing Cip10- P_{TET} -GTW-SP-DOC vector. Upon cloning and transformation in the toxin-sensitive *E. coli* DH5 α strain, we can select for vectors that successfully integrated a barcode sequence in replacement of the Doc toxin gene, improving the barcode recovery rate. The BC library described above was amplified using forward primer AVR11-Long-U1 and downstream primer BMT1-long-D1 using Phusion polymerase (NEB) (annealing temperature 60 °C). PCR products were then precipitated with ethanol and digested by AvrII and BmtI prior to ligation into AvrII and BmtI-digested Cip10- P_{TET} -GTW-SP-DOC using Rapid DNA ligase (Thermo Scientific). Ligation products were transformed into One Shot ccdB Survival™ 2 T1 *E. coli* competent cells (Invitrogen) and plated onto LB agar plates containing 100 mg ml⁻¹ ampicillin, incubated at 30 °C for 2 days. The 9000 transformation colonies were harvested by scraping from the plate surface and were grown in 150 ml 2 YT broth in triplicate flasks for 6 h at 30 °C. Cells were harvested and plasmid DNA extracted using Qiagen Maxi kit columns following manufacturer's recommendations. All plasmid DNA samples were pooled together, diluted 1:10 with EB buffer and then re-transformed into One Shot ccdB Survival™ 2 T1 *E. coli* competent cells (Invitrogen), transformed cells were then plated onto LB agar plates containing 100 mg ml⁻¹ ampicillin, incubated at 30 °C for 1–2 days. Individual colonies were selected, grown on 2 YT medium plus ampicillin overnight at 30 °C and stored in 25% glycerol prior to sending for sequencing of the barcodes.

2.4. Construction of *C. albicans* OE strains

Detailed methods for the transfer of *C. albicans* ORFs from pDONR207 into the expression vectors as well as the integration of the resulting expression plasmids at the *RPS1* locus have been described [21]. Briefly, ORFs were transferred from the entry clones into our collection of 2496 barcoded P_{TET} -driven overexpression vectors using the Gateway™ LR Clonase™ II Enzyme Mix (Invitrogen). After *E. coli* DH5 α transformation, the plasmids were verified by EcoRV digestion. The *URA3*-bearing expression plasmids were digested by *StuI* or *I-SceI* and transformed into *C. albicans*

strain CEC4642 (derived from SN76 [22]), according to Walther and Wendland [23], and adapted to 96-well microplate format as follows. From a 50 ml overnight YPD culture, 600 ml YPD were inoculated with CEC4642 at an OD600 of 0.2 and incubated at 30 °C to an OD600 of 0.8. The culture was then distributed in twelve 50 ml Falcon tubes and centrifuged at 3500 rpm for 5 min at 4 °C. Pelleted cells were washed with ice-cold 10X TE, resuspended in 1 ml TE 10:1/Lithium Acetate 100 mM (1X TELiOAc) and transferred to 2 ml Eppendorf tubes. After a 1 min centrifugation at 3500 rpm, cells were resuspended in 300 μ l 1X TELiOAc and kept on ice. The precipitated plasmid digests were transferred into deepwell 96 well plates with 5 μ l of salmon sperm DNA. 50 μ l of competent cells were added, along with 300 μ l of 4000 molecular weight polyethylene glycol (PEG₄₀₀₀) dissolved in 1 \times TELiOAc. After overnight incubation at 30 °C, the cells were exposed to 44 °C for 15 min followed by regeneration in 500 μ l SD + 0.4% glucose. The cells were then resuspended in 100 μ l SD + 0.4% glucose, plated onto SD agar plates and incubated at 30 °C for 3 days. Transformants were selected for prototrophy and verified by PCR using primers CipUL/CipUR that yields 1 kb product if integration of the OE plasmid had occurred at the *RPS1* locus.

2.5. Induction of the Tet-On system

Overexpression from P_{TET} was achieved by adding anhydrotetracycline (Atc, 3 μ g ml⁻¹; Fisher Bioblock Scientific) in YPD at 30 °C. Overexpression experiments were carried out in the dark, as Atc is light sensitive.

2.6. High-throughput screen for growth phenotypes on solid media

Strains in the P_{TET} -driven OE collection were grown in 96-well plate in YPD (30 h; 30 °C) and spotted on YPD plates supplemented or not with 3 μ g ml⁻¹ Atc using a 96-pin replicator. Plates were incubated at 30 °C for 2–5 days and photographed using the PhenoBooth from Singer Instruments at high quality (4128 \times 3096) or the Leica M80 stereomicroscope. Differences in colony size were automatically quantified using IRIS [24].

2.7. Preparation of strain pools

The 2454 barcoded OE strains, distributed in 32 96-well plates, were spotted on Nunc omnitray plates (Thermo Scientific) containing YPD agar supplemented with 50 μ g ml⁻¹ gentamycin using a 96-pin replicator and allowed to grow for 3 days at 30 °C. No major colony size alterations were noticed. Based on the number *n*

of spots on the plate, (0.05882 × n) milliliters of YPD supplemented with 50 µg ml⁻¹ gentamycin were added to each plate, and colonies were scraped off using a cell spreader. A complete pool was made with the 2454 OE strains. After adding glycerol at a final concentration of 15% and adjusting OD₆₀₀ to 50, cells were aliquoted in 1-ml tubes and frozen at -80 °C.

2.8. Barcode sequencing

To quantify the relative concentration of each barcoded strain in mixed populations, genomic DNA was prepared from 1 ml of OD₆₀₀ = 50 culture, using the MasterPure Yeast DNA Purification Kit (Epicentre). Libraries were prepared following the 16S Metagenomic Sequencing Library Preparation protocol (Illumina, Inc., San Diego, United States) recommendations. In addition to the so-called Illumina overhang sequences, spacer bases (0, 4, 8, and 12 N) were added to the forward primers to shift the signal during sequencing to avoid mistakes during the sequencing due to the similar flanking regions of the BC amplicons. The purified amplicons, which contained the Illumina overhang, were then indexed with Illumina

Nextera XT Index kit (Illumina, Inc., San Diego, CA, USA). Libraries were quantified with Qubit and then quality-checked and quantified on a Bioanalyzer (Agilent). The libraries were pooled in equimolar amounts and sequenced on an Illumina NextSeq500 Sequencing System with 75 bp single-reads. Raw data have been deposited in the NCBI Sequence Read Archive under BioProject ID PRJNA911593.

For barcode quantification in the populations from the pool enrichment assay, sequence read files were aligned against the collection of barcodes using blastn (blast + v 2.10.1). To determine the occurrence of each barcode for each time point and condition, each fastq file was first converted into FASTA format to be used as query and the barcode collection was transformed beforehand with makeblastdb to be used as database. Upon blastn search, counts for each time point and condition were merged into a single text file.

To study the contribution of treatment and biological variations to Bar-seq data, we applied the negative binomial models used to analyze RNA-seq data. RNA-seq analysis tools such as DESeq2 [25] are readily adapted to Bar-seq analysis as previously demonstrated [26], providing estimates of effect sizes and their statistical

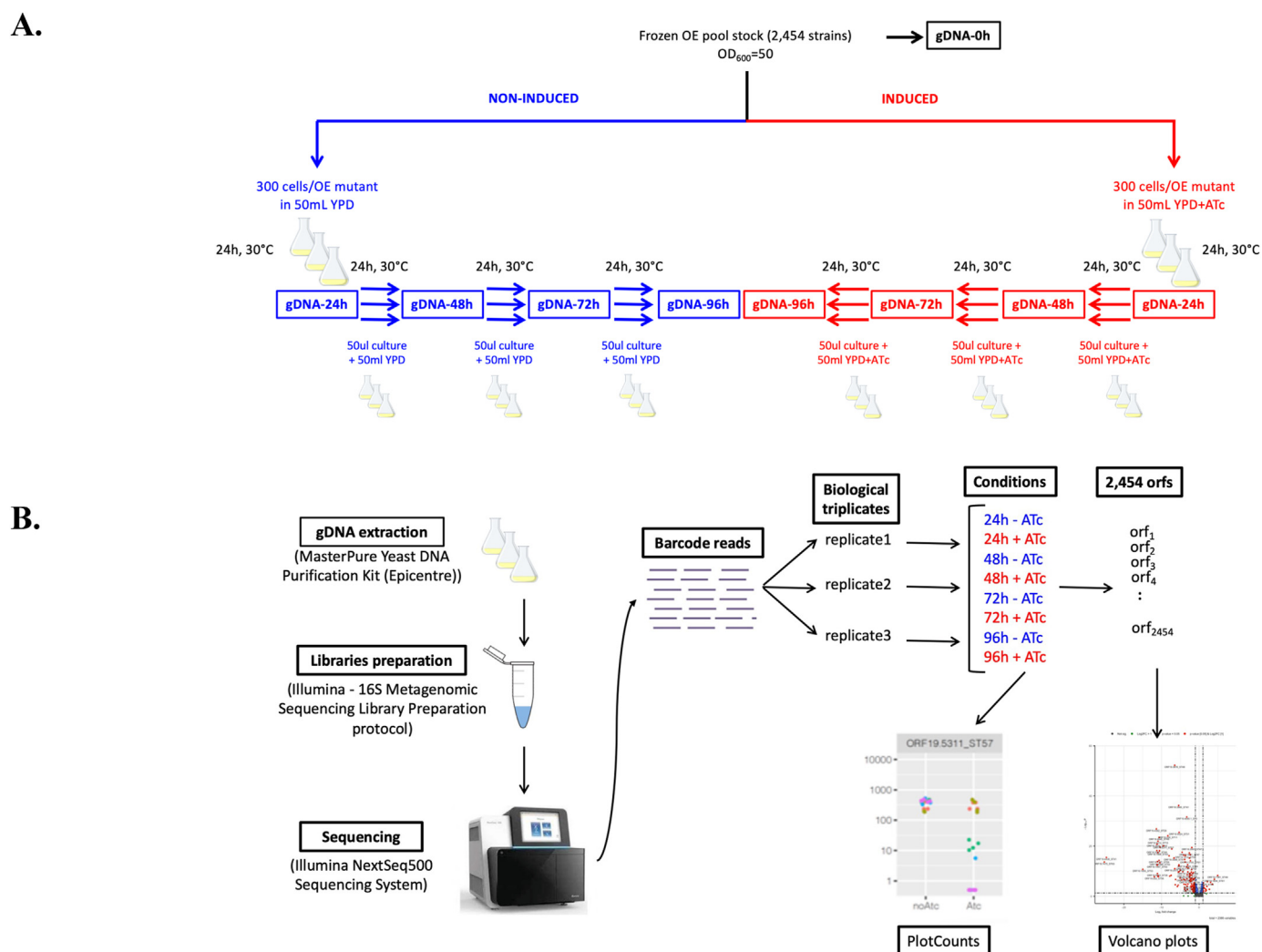


Fig. 1. Schematic representation of experimental set up for the pooled growth assay and barcode detection (A). Timeline for the pooled growth assay. Cultures were inoculated from the frozen pool. Cells were then grown for 96 h. Every 24 h, cells were harvested for further genomic DNA extraction and passaged in fresh media (B). Barcode detection. Genomic DNA was purified from the cell pellets using the MasterPure Yeast DNA Purification Kit (Epicentre). Barcodes were amplified from genomic DNA and subjected to sequencing using an Illumina NextSeq500 sequencing system. Sequencing data were then analyzed for barcode counting in the different conditions. The abundance of each OE mutant can be represented in PlotCounts to compare variation across the different growth conditions for a given ORF or in Volcano plots to compare the abundance of the OE mutants in a given growth condition.

significance for each time point and conditions. Proportion of each strain in the population was subsequently calculated and plotted using DESeq2 library under Open Source R Software (v 4.0.3). Volcano plots are plotted using EnhancedVolcano library (10.18129/B9.bioc.EnhancedVolcano). At the initial time point (0 h in Fig. 1), only 3% of the ORFs are not recovered upon the bioinformatics treatment of read sequences.

2.9. *In vitro* competitive fitness assay

The pool frozen at OD600 of 50 was diluted to an OD600 of 5 and a 340 µl aliquot was used to inoculate 50 ml of YPD medium ± Atc, allowing an inoculum with ~300 cells per strain. The cultures were then incubated at 30 °C with agitation (200 rpm) for 96 h, in the absence or presence of Atc. Every 24 h, 50 µl were used for re-inoculating 50 ml of fresh YPD and 50 OD600-unit cultures were used for DNA extraction and barcode sequencing.

2.10. Confirmation of the *in vitro* fitness-profiling data by individual liquid growth assay

C. albicans strains were individually grown three times independently in 96-well microtiter plates in YPD, either with or

without Atc. Cells were inoculated at a starting OD600 = 0.01 in a final volume of 100 µl. The yeast strains were grown, shaking, for 48 h at 30 °C and the OD600 was monitored every 10 min using a microplate reader (TECAN Sunrise). Doubling-time (DT) was calculated by dividing by 2 the time between OD600 = 0.15 and OD600 = 0.6.

3. Results and discussion

3.1. Development of a collection of barcoded overexpression vectors

We previously reported the establishment of a *C. albicans* ORFeome where 5099 ORFs were successfully cloned and full-length sequence-validated [4]. The entire entry clone collection, representing 5099 unique genes, is arrayed in 64 96-well plates numbered BP1 to BP64. The *C. albicans* ORFeome was developed in pDONR207 as this allows subsequent transfer of the cloned ORFs to Gateway-adapted destination vectors [27]. We leveraged the ORFeome to develop a collection of Gateway vectors for inducible OE in *C. albicans*. To increase experimental throughput, we aimed at generating a collection of barcoded OE plasmids, meaning that a unique 25bp randomized sequence (barcode) serves as a unique identifier for OE plasmids and subsequently *C. albicans* OE strains (Fig. 2).

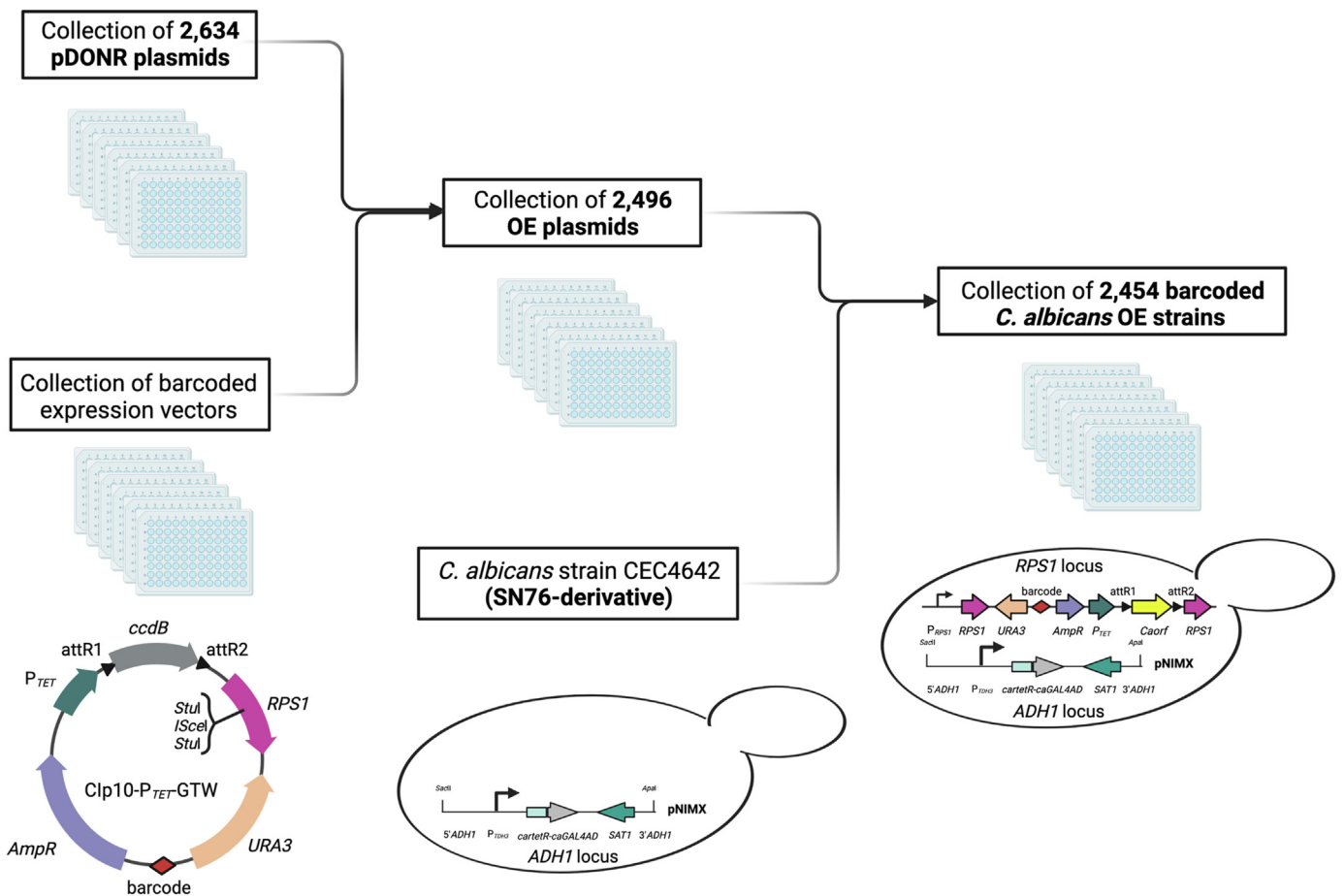


Fig. 2. Construction of the largest collection of *C. albicans* OE strains. ORFs in the **2634 pDONR plasmids** are individually transferred into a barcoded destination vector Clp10-P_{TET}-GTW, using the Gateway system. The schematic map of the Clp10-P_{TET}-GTW vector is shown. The *URA3* gene is used for selection of *C. albicans* transformants and derivatives of these OE vectors can be targeted to the *C. albicans RPS1* locus when linearized with *StuI* or *I-SceI*. This high-throughput recombinational cloning of the ORFs resulted in a collection of **2496 barcoded OE plasmids**. The *StuI* or *I-SceI*-digested OE plasmids are transformed into the recipient strain resulting in a collection of **2454 barcoded *C. albicans* OE strains**. The *C. albicans* recipient strain is a derivative of the reference strain SC5314 [22] and has been modified to constitutively express the tetracycline-dependent transactivator rtTA, allowing transcriptional activation at the *TET* promoter.

3.1.1. The recipient overexpression vector for the barcodes

The Cip10- P_{TET} -GTW vector [5] was modified by inserting the SP cassette at the SacII site. The SP cassette contains the SP1 and SP2 sequencing primers from the Illumina technology with in between (i) a 17bp fragment that carries the *AvrII* and *BmtI* restriction sites and that will serve as the landing platform for the barcodes as well as (ii) a fragment of the Kanamycin resistance gene to act as a spacer and allow proper sequencing of the barcodes by Illumina sequencing technology with the SP1 and SP2 primers. Using a P_{TET} -*gLuc59* reporter system [5], we first confirmed that the insertion of the SP cassette at the SacII site in the Cip10- P_{TET} -GTW vector [4] did not alter the level of P_{TET} -driven expression (data not shown). To optimize barcode cloning in the Cip10- P_{TET} -GTW-SP vector, we used the *phd/doc* toxin-antitoxin modules [20]. The *Doc* toxin gene was ligated within the 17bp landing platform of Cip10- P_{TET} -GTW-SP and transformation in *E. coli* ECC1081 strain harboring the gene encoding for the PHD antitoxin at the *attTn7* locus (*glmS::PHD*) allowed obtaining the Cip10- P_{TET} -GTW-SP-DOC vector. Upon cloning and transformation in the toxin-sensitive *E. coli* DH5 α strain, we can select for vectors that successfully integrated a barcode sequence in replacement of the *Doc* toxin gene, improving the barcode recovery rate (Fig. S1).

3.1.2. Generation of the barcoded vector library

We have generated several python scripts to filter out 65,777 25mer DNA barcode sequences from the list designed by Xu et al. [14] according to selected criteria as described below. We excluded the DNA sequences containing restriction enzyme sites that were to be used for subsequent cloning of the library, as well as sequences that displayed high similarities to the *C. albicans* genome or cloning vectors, resulting in 65,375 validated barcodes. The oligonucleotides, synthesized on Agilent microarrays, were designed such that each 25mer barcode sequences was flanked by common primer sequences U1 and D1, and were cloned in the recipient overexpression vector Cip10- P_{TET} -GTW-SP-DOC.

In this study, we have taken advantage of the Gateway methodology to transfer all individual ORFs contained in odd number ORFeome plates (2634 ORFs) into an array of destination vectors each containing a unique barcode, resulting in a collection of 2496 P_{TET} -driven OE plasmids, each containing a *C. albicans* ORF associated to a unique 25bp barcode (Table S1).

3.2. Establishment of a collection of *C. albicans* OE strains

The recipient strain for the collection of OE plasmids is CEC4642, a derivative of *C. albicans* strain SN76 [22] (an auxotroph derivative of the reference strain SC5314 [28]) transformed with the linearized plasmid pNIMX [5] that provides the *C. albicans*-adapted reverse Tet-dependent transactivator (*cartTA*) required for the activation of the P_{TET} promoter in a tetracycline-dependent manner. Also, to optimize subsequent *in vivo* experiments, arginine and histidine prototrophies have been restored in CEC4642. To ensure that the successive transformation steps to obtain CEC4642 did not trigger any phenotype, we verified its filamentation properties in both liquid and solid media, as well as its growth rate in rich and poor media. We did not observe any changes when compared to the lab reference strain SC5314 (Figs. S2A and B).

We then investigated the ability of the recipient strain CEC4642 to induce OE in the proper conditions. To do so, we introduced the *StuI*-linearized Cip- P_{TET} -*UME6* OE plasmid [4] in CEC4642. *UME6* encodes a transcription factor whose OE has been shown to trigger filamentation [29,30]. As expected, we were able to trigger *C. albicans* filamentation when CEC4642 strains transformed with a *UME6* OE plasmid were grown in the presence of anhydrotetracycline, validating the use of CEC4642 as a recipient strain for the OE plasmid collection (Fig. S2C).

Prior to transforming CEC4642 with the entire OE plasmid collection, we selected 12 random barcoded vectors to be tested with *UME6* to generate 12 expression plasmids. We were able to successfully transfer the *UME6* ORF into the 12 barcoded OE vectors.

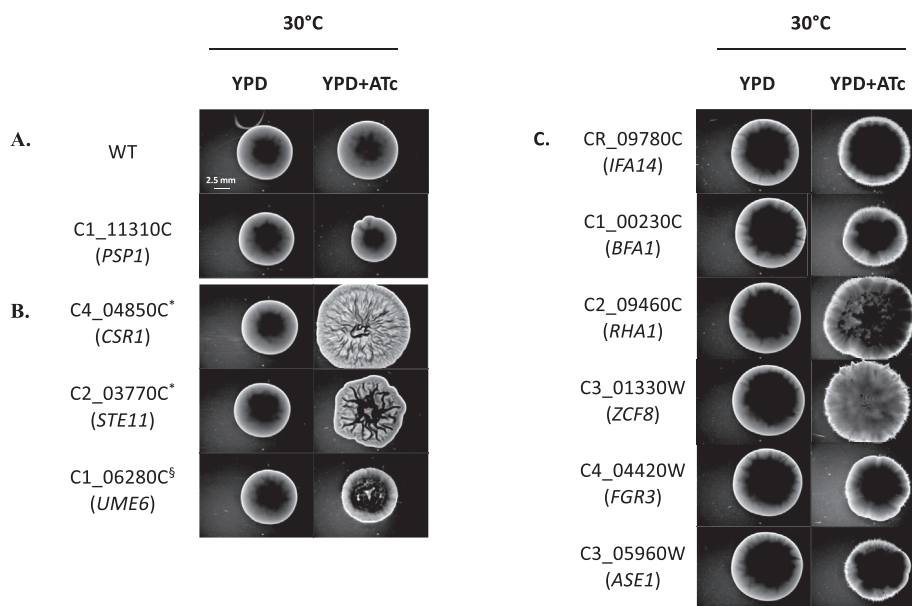


Fig. 3. Growth and colony morphology phenotypes of *C. albicans* OE strains. (A) Example of an OE mutant showing growth defect upon OE along with the parental strain CEC4642 (SN76 *ARG4 HIS1 URA3*) (B) and (C). OE mutants with enhanced colony wrinkling upon OE on YPD + ATc ($3 \mu\text{g ml}^{-1}$) at 30 °C. Genes labelled * were already identified as regulators of morphogenesis and fitness in our previous screen [5] and genes labelled § were already known to be involved in morphogenesis regulation [31,32]. Colonies were photographed after 4 days (panels A and B) or 6 days (panel C) of growth on YPD \pm ATc at 30 °C using a Leica M80 stereomicroscope. Scale bar = 2.5 mm. All pictures are at the same scale.

The *UME6* OE plasmids were then successfully transformed in CEC4642 and shown to integrate in the *RPS1* locus as expected. Further, the resulting transformed strains were shown to filament upon anhydrotetracycline addition (Fig. S2C).

From there, all barcoded P_{TET} -driven OE plasmids from the collection were individually transformed into the *C. albicans* recipient strain CEC4642, resulting in a collection of 2454 OE mutants, which constitutes, to the best of our knowledge, the largest OE collection in *C. albicans*.

3.3. Individual screening for genes affecting morphogenesis or fitness upon OE

To exemplify the utility of the OE mutants, we have undertaken an extensive OE screen to identify mutants with morphological or growth phenotypes upon spotting on solid media containing anhydrotetracycline. Out of 2454 OE mutants, we identified 51 genes whose OE results either in (i) growth defects, as illustrated by smaller spots on YPD + ATc agar plates as compared to YPD agar plates or (ii) colony morphological phenotypes, as illustrated by the unusual appearance of the colony on YPD + ATc agar plates as compared to YPD plates (Fig. 3, Table S2). Eight out of these 51 OE mutants display a growth defect (1 example is illustrated in Fig. 3A) and 43 display a colony morphology phenotype (9 examples are illustrated in Fig. 3B and C). Among these 51 mutants, we found six genes that were already identified as regulators of morphogenesis and fitness in our previous screen [5], as well as other genes known to be involved in the regulation of morphogenesis such as *UME6* [31] (Fig. 3B) and *PHO81* [32].

3.4. Utilization of barcoded overexpression strains for pool experiments

We employed the pooled *C. albicans* OE collection to examine the impact of gene OE on the physiology of *C. albicans*. The pool of 2454 OE strains was grown in liquid YPD at 30 °C in the absence and

presence of ATc. Cultures were diluted fresh at successive time-points (24, 48, 72, and 96 h) to maintain exponential growth (Fig. 1A). At each timepoint, genomic DNA was extracted, barcodes were amplified, and sequencing libraries were prepared as described in M&M (Fig. 1B). Sequencing was performed on an Illumina NextSeq 500. We looked for OE mutants that exhibit differences in abundance because of ATc treatment, illustrated by changes in the number of BC reads. The validated mutants were identified based on the fold-change (FC) in absence and presence of ATc, as well as adjusted p-value ($|\log_2(\text{FC})| > 1$ and $\text{padj} < 0.0001$). Using these criteria, we identified 65, 122 and 176 genes whose OE results in a change in fitness at 48 h, 72 h and 96 h respectively (Figs. 4A and S3, Table S2). Eight genes were identified at 24 h as having a $|\log_2(\text{FC})| > 1$ but no significant padj . Overall, the majority of the OE mutants exhibiting a phenotype are less fit than the wild-type. The 193 unique genes (Table S2) identified in these screens, both on solid media and in pool liquid experiments, were biased for genes with particular GO biological process or molecular function annotations, such as genes involved in different aspects of filamentous growth (Fig. 5).

3.5. Validation of fitness estimated in overexpression mutant pools with targeted overexpression strains

To determine whether the growth deficiencies or advantages of OE mutant strains in the pooled assays were predictive of the mutant behavior when grown in isolation, we used liquid growth assays of individually grown strains. The experimental design for fitness analysis of individual OE mutants led us to compare the fitness of 91 strains measured in the pooled assays at 48 h to that measured in individual fitness assays; **49/91** had a $\text{Log}_2\text{FC} < -1$ (fitness loss), **9/91** had a $\text{Log}_2\text{FC} > 1$ (fitness gain) and for **33/91**, $-1 < \text{Log}_2\text{FC} < 1$. Correlations calculated across all samples using the Pearson's correlation coefficient ($\text{PCC} = 0.7560$) show a strong positive correlation between the values obtained via the two methods (Fig. 4B).

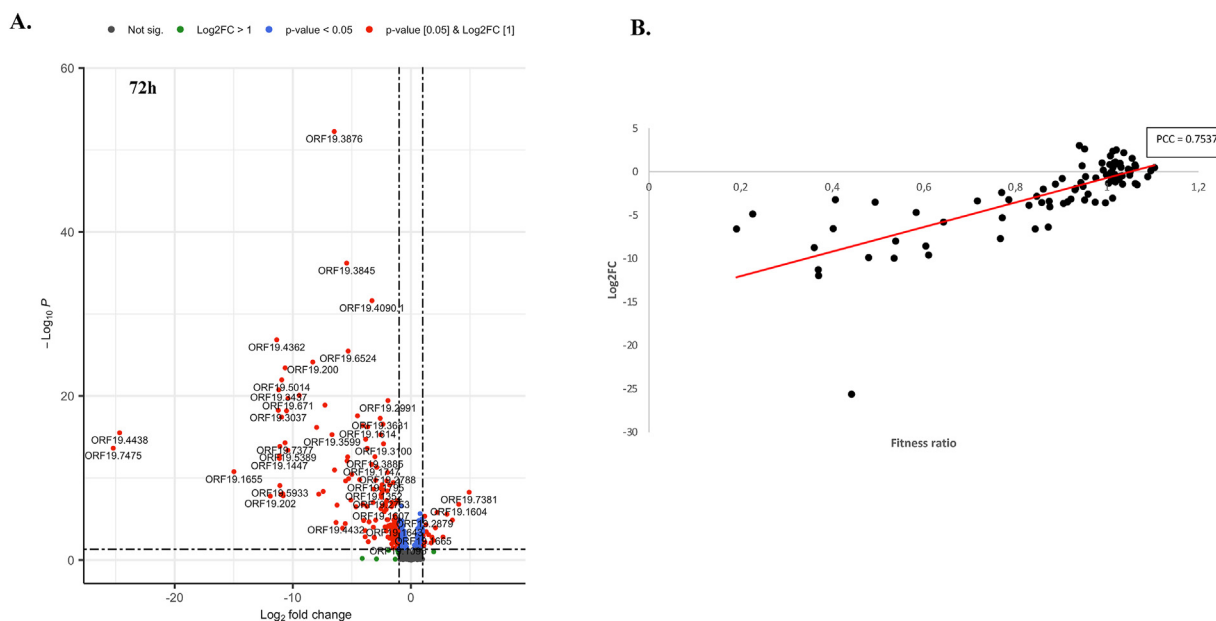


Fig. 4. Impact of induced-OE on fitness (A). Volcano plot of OE mutant abundance in presence of ATc relative to their abundance in absence of ATc at 72 h (\log_2 fold-change), based on Bar-seq experiment. Significance was determined using a fold-change (FC) cut-off of $|\log_2(\text{FC})| > 1$ and an adjusted p-value (padj) cut-off of < 0.0001 (B). Scatter plot showing the agreement of fitness estimates between OE mutant pools (Bar-seq) and individual OE strain for 91 mutants. The \log_2 fold-change (FC) of barcode abundance (+ATc relative to -ATc) is plotted against the fitness ratio defined as (doubling time in -ATc)/(doubling time in +ATc) of individual OE mutants. Red: linear regression between $\log_2\text{FC}$ and fitness ratio, along with Pearson correlation coefficient (PCC). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

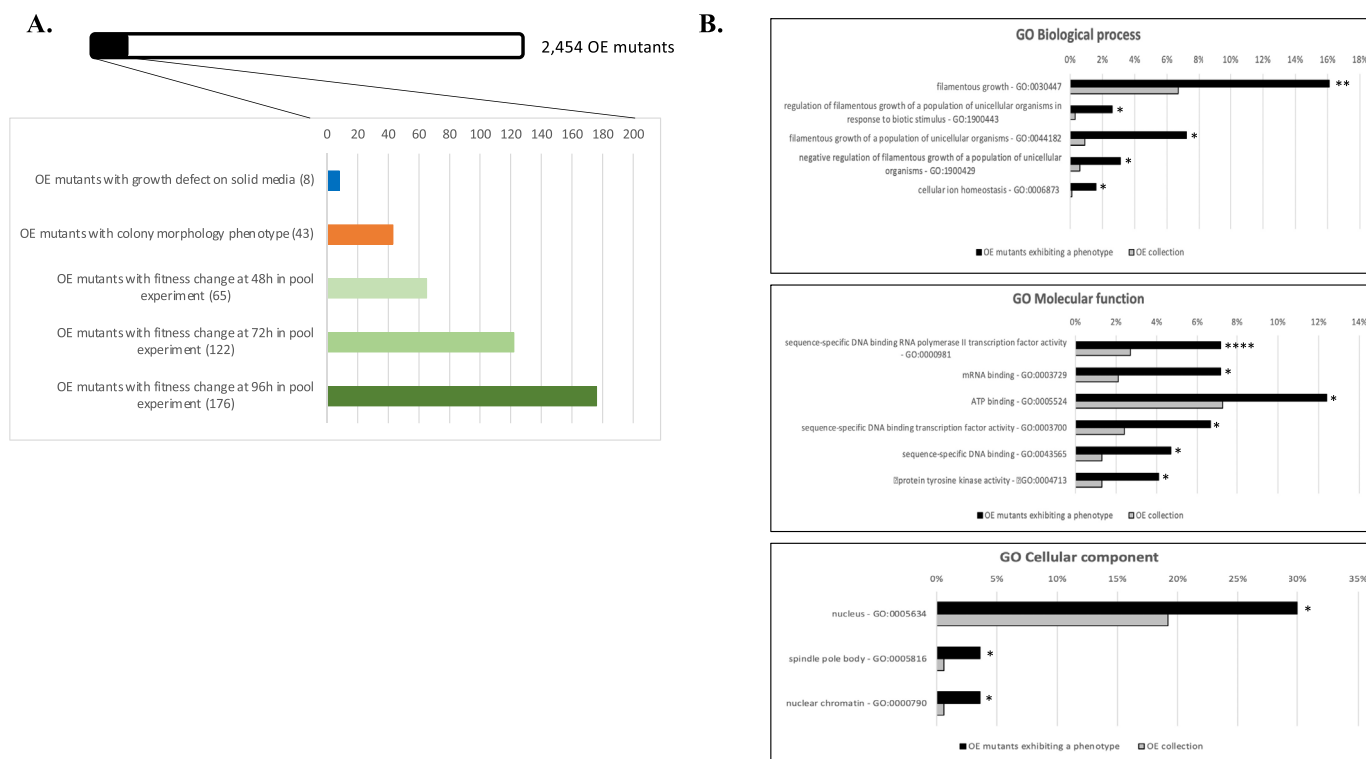


Fig. 5. Schematic representation of the *C. albicans* overexpression collection (A). The top bar represents the 2454 genes present in the collection. The blue bar represents the **eight genes** whose OE caused slow growth on solid media; the orange bar represents the genes whose OE caused an aberrant morphological phenotype upon growth on solid media (**42**). The genes whose OE results in fitness changes in pool experiment are depicted by the green bars: fitness change at 48 h, light green (**65**); fitness change at 72 h, medium green (**122**); fitness change at 96 h, dark green (**176**) (B). Enrichment within the fitness and morphology gene datasets for biological process, molecular function and cellular component. Genes in our dataset were classified by FungiFun (<https://elbe.hki-jena.de/fungifun/fungifun.php>). Enrichment in the various categories relative to the collection of 2454 OE mutants - (grey columns) is depicted. (* $\text{padj} < 0.05$, ** $\text{padj} < 0.01$, *** $\text{padj} < 0.001$, **** $\text{padj} < 0.0001$). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4. Conclusions

Even with the advent of *C. albicans*-adapted CRISPR Cas9 technology, gene deletion in *C. albicans* remains a tedious process, mainly due to its heterozygous diploid genome and the relatively low efficiency of gene replacement approaches. To circumvent this, various OE collections have been generated and screens have been performed to gain insights into different aspects of *C. albicans* biology, such as morphological transitions, biofilm formation, genome stability, drug resistance and gut colonization [7,33]. We previously established an ORFeome collection (CandidaOrfDB database at <http://candidaorfeome.eu>) covering 83% of *C. albicans* genome with the aim to enable genome- and proteome-wide systematic analysis of *C. albicans* to improve our understanding of the molecular networks and mechanisms underlying and governing its pathogenesis. By generating a collection of 2496 P_{TET} -driven bar-coded OE plasmids, and a collection of 2454 *C. albicans* bar-coded OE mutant strains, we are expanding the use of the ORFeome and validating the use of systematic OE approaches to study *C. albicans*.

In a fungal human pathogen where 69% of the ORFs remain to be characterized, clustering approaches based on systematic phenotypic characterization of our collection of OE mutants will allow inferring functions of uncharacterized genes or at least assign them to a cellular process or pathway. Among these genes, it is likely that a subset will be fungal specific and therefore could represent potential targets for drug discovery.

Importantly, the OE collections presented here constitute versatile tools that have several applications in WT and mutant backgrounds. Indeed, genetic interactions can be inferred from more sophisticated suppressor or enhancer screens that determine

whether OE of certain genes can complement the genetic or chemical blockade of specific pathways, or on the contrary cause a chemical-than additive phenotype. At the inter-species scale, OE of *C. albicans* genes in heterologous organisms could functionally complement recessive mutations or in contrast cause deleterious effects in the host and provide insights in gene function.

While we present here a collection of 2496 OE plasmids (corresponding to the odd-numbered BP plates [4]), the rest of the ORFeome (corresponding to the even-numbered BP plates [4]) is already being transferred into bar-coded P_{TET} -driven OE vectors and OE strains are already being generated. Soon, the nearly complete collection of OE mutants, expected to cover >80% of *C. albicans* genome will allow large-scale, high-throughput phenotyping experiments to fuel “phenomics” projects and define biological networks in *C. albicans*.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgements

This work was financially supported by a Biomedical Resources grant from the Wellcome Trust (The *Candida albicans* ORFeome project, 088858/Z/09/Z to CAM and CD), the European Commission (FINSysB, PITN-GA-2008-214004 to CD), and the Agence Nationale de la Recherche (KANJI, ANR-08-MIE-033-01 to CD; CANDICOL, ANR-10-01 to CD). CAM would like to acknowledge support from a Royal Society Leverhulme Trust Senior Research Fellowship (SRF\R1\180115). We acknowledge support from the French

Government's Investissement d'Avenir program (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases, ANR-10-LABX-62-IBEID).

We are grateful to other members of our groups for their support throughout the development of the *C. albicans* ORFeome.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.resmic.2022.104025>.

References

- [1] Brown GD, Denning DW, Levitz SM. Tackling human fungal infections. *Science* 2012;336:647. <https://doi.org/10.1126/science.1222236>.
- [2] Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. *Sci Transl Med* 2012;4:165rv13. <https://doi.org/10.1126/scitranslmed.3004404>.
- [3] Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G. The *Candida* Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res* 2017;45:D592–6. <https://doi.org/10.1093/nar/gkw924>.
- [4] Legrand M, Bachellier-Bassi S, Lee KK, Chaudhari Y, Tournu H, Arbogast L, et al. Generating genomic platforms to study *Candida albicans* pathogenesis. *Nucleic Acids Res* 2018;46:6935–49. <https://doi.org/10.1093/nar/gky594>.
- [5] Chauvel M, Nesseir A, Cabral V, Znaidi S, Goyard S, Bachellier-Bassi S, et al. A versatile overexpression strategy in the pathogenic yeast *Candida albicans*: identification of regulators of morphogenesis and fitness. *PLoS One* 2012;7:e45912. <https://doi.org/10.1371/journal.pone.0045912>.
- [6] Loll-Kripplleber R, Feri A, Nguyen M, Maufrais C, Yansouni J, d'Enfert C, et al. A FACS-optimized screen identifies regulators of genome stability in *Candida albicans*. *Eukaryot Cell* 2015;14:311–22. <https://doi.org/10.1128/ec.00286-14>.
- [7] Jaitly P, Legrand M, Das A, Patel T, Chauvel M, Maufrais C, et al. A phylogenetically-restricted essential cell cycle progression factor in the human pathogen *Candida albicans*. *Nat Commun* 2022;13:4256. <https://doi.org/10.1038/s41467-022-31980-3>.
- [8] Mazurkiewicz P, Tang CM, Boone C, Holden DW. Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. *Nat Rev Genet* 2006;7:929–39. <https://doi.org/10.1038/nrg1984>.
- [9] Oh J, Nislow C. Strain engineering, methods and protocols. *Methods Mol Biol* 2011;765:225–52. https://doi.org/10.1007/978-1-61779-197-0_14.
- [10] Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, et al. Targeted changes of the cell wall proteome influence *Candida albicans* ability to form single- and multi-strain biofilms. *PLoS Pathog* 2014;10:e1004542. <https://doi.org/10.1371/journal.ppat.1004542>.
- [11] Znaidi S, Wijlick L, Hernández-Cervantes A, Sertour N, Desseyn J, Vincent F, et al. Systematic gene overexpression in *Candida albicans* identifies a regulator of early adaptation to the mammalian gut. *Cell Microbiol* 2018;20:e12890. <https://doi.org/10.1111/cmi.12890>.
- [12] Delarze E, Brandt L, Trachsel E, Patxot M, Pralong C, Maranzano F, et al. Identification and characterization of mediators of fluconazole tolerance in *Candida albicans*. *Front Microbiol* 2020;11:591140. <https://doi.org/10.3389/fmicb.2020.591140>.
- [13] Xu Q, Schlabach MR, Hannon GJ, Elledge SJ. Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc Natl Acad Sci* 2009;106:2289–94. <https://doi.org/10.1073/pnas.0812506106>.
- [14] Muzzey D, Schwartz K, Weissman JS, Sherlock G. Assembly of a phased diploid *Candida albicans* genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. *Genome Biol* 2013;14:R97. <https://doi.org/10.1186/gb-2013-14-9-r97>.
- [15] Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, Munro CA, et al. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 2009;459:657–62. <https://doi.org/10.1038/nature08064>.
- [16] Dujon B, Sherman D, Fischer G, Durrrens P, Casaregola S, Lafontaine I, et al. Genome evolution in yeasts. *Nature* 2004;430:35–44. <https://doi.org/10.1038/nature02579>.
- [17] Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 2002;418:387–91. <https://doi.org/10.1038/nature00935>.
- [18] Watkins KH, Stewart A, Fairbrother W. A Rapid high-throughput method for mapping ribonucleoproteins (RNPs) on human pre-mRNA. *Jove* 2009. <https://doi.org/10.3791/1622>.
- [19] Guérout A-M, Iqbal N, Mine N, Ducos-Galand M, Melderer LV, Mazel D. Characterization of the phd-doc and ccd toxin-antitoxin cassettes from *Vibrio* superintegrons. *J Bacteriol* 2013;195:2270–83. <https://doi.org/10.1128/jb.01389-12>.
- [20] Cabral V, Chauvel M, Firon A, Legrand M, Nesseir A, Bachellier-Bassi S, et al. Host-fungus interactions, methods and protocols. *Methods Mol Biol* 2012;845:227–44. https://doi.org/10.1007/978-1-61779-539-8_15.
- [21] Noble SM, Johnson AD. Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. *Eukaryot Cell* 2005;4:298–309. <https://doi.org/10.1128/ec.4.2.298-309.2005>.
- [22] Walthers A, Wendland J. PCR-based gene targeting in *Candida albicans*. *Nat Protoc* 2008;3:1414–21. <https://doi.org/10.1038/nprot.2008.137>.
- [23] Kritikos G, Banzhaf M, Herrera-Dominguez L, Koumoutsis A, Wartel M, Zietek M, et al. A tool named Iris for versatile high-throughput phenotyping in microorganisms. *Nat Microbiol* 2017;2:17014. <https://doi.org/10.1038/nmicrobiol.2017.14>.
- [24] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;11:R106. <https://doi.org/10.1186/gb-2010-11-10-r106>.
- [25] Robinson DG, Chen W, Storey JD, Gresham D. Design and analysis of bar-seq experiments. *G3 Genes Genomes Genetics* 2013;4:11–8. <https://doi.org/10.1534/g3.113.008565>.
- [26] Walhout AJ, Temple GF, Brasch MA, Hartley JL, Lorz MA, Heuvel S van den, et al. [34] GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* 2000;328:575–IN7. [https://doi.org/10.1016/s0076-6879\(00\)28419-x](https://doi.org/10.1016/s0076-6879(00)28419-x).
- [27] Gillum AM, Tsay EYH, Kirsch DR. Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF* mutations. *Mol Gen Genet* 1984;198:179–82. <https://doi.org/10.1007/bf00328721>.
- [28] Carlisle PL, Banerjee M, Lazzell A, Monteagudo C, López-Ribot JL, Kadosh D. Expression levels of a filament-specific transcriptional regulator are sufficient to determine *Candida albicans* morphology and virulence. *Proc Natl Acad Sci USA* 2009;106:599–604. <https://doi.org/10.1073/pnas.0804061106>.
- [29] Zeidler U, Lettner T, Lassnig C, Müller M, Lajko R, Hintner H, et al. *UME6* is a crucial downstream target of other transcriptional regulators of true hyphal development in *Candida albicans*. *Fems Yeast Res* 2009;9:126–42. <https://doi.org/10.1111/j.1567-1364.2008.00459.x>.
- [30] Banerjee M, Thompson DS, Lazzell A, Carlisle PL, Pierce C, Monteagudo C, et al. *UME6*, a novel filament-specific regulator of *Candida albicans* hyphal extension and virulence. *Mol Biol Cell* 2008;19:1354–65. <https://doi.org/10.1091/mbc.e07-11-1110>.
- [31] Chung S-C, Kim T-I, Ahn C-H, Shin J, Oh K-B. *Candida albicans* *PHO81* is required for the inhibition of hyphal development by farnesic acid. *Febs Lett* 2010;584:4639–45. <https://doi.org/10.1016/j.febslet.2010.10.026>.
- [32] Rai LS, Wijlick L, Chauvel M, d'Enfert C, Legrand M, Bachellier-Bassi S. Overexpression approaches to advance understanding of *Candida albicans*. *Mol Microbiol* 2022;117:589–99. <https://doi.org/10.1111/mmi.14818>.